A New Criterion for Nonlinearity of Block Ciphers

Orr Dunkelman
Katholieke Universitiet Leuven,
Department of Electrical Engineering ESAT/SCD-COSIC,
Kasteelpark Arenberg 10, Heverlee-Leuven B-3001, Belgium.
Email: orr.dunkelman@esat.kuleuven.be
Nathan Keller
Einstein Institute of Mathematics,
Hebrew University,
Jerusalem 91940, Israel.
Email: nkeller@math.huji.ac.il

Abstract— **For years, the cryptographic community has searched for good nonlinear functions. Bent functions, almost perfect nonlinear functions, and similar constructions have been suggested as a good base for cryptographic applications due to their highly nonlinear nature. In the first part of this paper we examine using these functions as block ciphers, and present several distinguishers between almost perfect nonlinear permutations and random permutations. The data complexity of the best distinguisher is $\Theta(2^{4n/9})$ and its time complexity is $\Theta(2^{4n/9})$ for an n-bit block size, independent of the key size.**

In the second part of the paper we suggest a criterion to measure the effective linearity of a given block cipher. We devise a general distinguisher for block ciphers based on their effective linearity. Finally, we show that for several constructions, our distinguishing attack is better than previously known techniques.

Index Terms— **Almost perfect nonlinear permutations, highly nonlinear functions, effective linearity, differential cryptanalysis**

I. INTRODUCTION

For years, highly nonlinear functions were extensively used in various cryptographic applications. Highly nonlinear functions were promoted since ciphers that are close to linear are susceptible to various approximation attacks. *Differential cryptanalysis* [5] and *linear cryptanalysis* [12] show that even partial approximations of the encryption algorithm by a linear function are sufficient to mount powerful distinguishing and key-recovery attacks.

In [13] it was suggested to use (almost) perfect nonlinear functions (functions with a maximal distance from all linear structures) in block ciphers. The nonlinear functions can be used either as a building block in the structure of the block cipher (such as an S-box [16], [17], [18]) or as the entire block cipher (as discussed in [1]). Such (almost) perfect nonlinear constructions can be used to prove security against differential and linear cryptanalysis [19].

In [16] the following construction, later named in [19], was suggested: An almost perfect nonlinear permutation (APNP) is a permutation $f : GF(2^n) \to GF(2^n)$ such that for any $a \neq 0$, the function $g(x, a) = f(x) \oplus f(x \oplus a)$ assumes exactly 2^{n-1} different values. In an almost perfect nonlinear permutation, for any two pairs of distinct input values with the same difference, the corresponding output pairs do not have the same difference. Due to this property, APNPs are considered secure against differential cryptanalysis [5] as well as against linear cryptanalysis [12].[1] A permutation that is very close to be an APNP is the S-box SubBytes of AES [8]. In the SubBytes permutation, for a non-zero input difference α, there are two pairs of inputs with difference α and the same output difference, while the other 126 pairs with input difference α have different output differences.

In the first part of this paper we analyze the concept of using highly nonlinear functions (such as APNPs) as the entire block cipher. We devise several distinguishers between an APNP and a random permutation[2] based on the "too good" differential properties of APNPs. The main result of this part of the paper is the following:

Theorem 1: Let $f : GF(2^n) \to GF(2^n)$ be a black box permutation for which we have to determine whether it is an APNP or a random permutation, and let $D = 1.794 \cdot 2^{n/3+l/4}$ where $l \geq 0$. There exists an algorithm with data complexity of D chosen plaintexts and time complexity of $D + D^2 \cdot 2^{-l-1}$ memory accesses, such that:

1) If f is an APNP then the algorithm outputs "APNP" with probability 1.
2) If f is a random permutation, the algorithm outputs "Random Permutation" with probability $p > 0.8$.

In particular (substituting $l = 4n/9$), there exists a distin-

[1] Note that there are two definitions to the term nonlinearity — nonlinearity as distance from linear functions, and nonlinearity in the sense of almost perfect nonlinear permutations. In this paper we use the latter definition.

[2] We note that an ideal block cipher can be modeled as a random permutation, i.e., a permutation picked randomly from all the permutations over $GF(2^n)$, where n is the block size. A random permutation is considered to be the "ideal cipher" as for such a permutation the knowledge of all but few plaintext/ciphertext pairs does not shed light on the last few, i.e., given a plaintext which is not in the known set, each of the ciphertexts which is not in the set might be its encryption. However, in order to describe a random permutation over $GF(2^n)$ one has to store $\log_2(2^n!) = O(n \cdot 2^n)$ bits as a secret key, and hence using truly random permutations for encryption is computationally infeasible.

A preliminary version of this paper was presented in RSA Conference 2006, Cryptographers' Track.

The first author was partially supported by the Clore scholarship programme. Some of the work described in this paper was done while the author was in the computer science department of the Technion in Israel.

The second author was partially supported by the Adams Fellowship.

guisher between an APNP and a random permutation with data and time complexities of $\Theta(2^{4n/9})$.

The distinguishers are then extended to ciphers with close to uniform difference distribution tables (that is, ciphers that are close to APNPs). This result leads to the conclusion that despite their favorable properties, highly nonlinear ciphers might possess inherent weakness.

In a way, the concept of this part of the paper is similar to the concept behind impossible differential cryptanalysis [4]. Usually, differential cryptanalysis uses differentials with high probability. The idea of impossible differential attacks is to exploit differentials with zero probability. The attacker utilizes the fact that the differential properties are "too strong". In the same way, our attacks recognize APNPs due to their "too high" level of nonlinearity.

In the second part of this paper we analyze the following encryption scheme: Let f be a permutation, f^{-1} be its inverse, and K be a randomly chosen non-zero key. We define $g(x) = f^{-1}(f(x) \oplus K)$, and show that g has several predictable differential properties. Those properties are closely related to the level of nonlinearity of the function f, and in particular to the sum of squares of the entries in the differential distribution table[3] of f. We use those properties to define the *effective linearity* of the function f, and show that the effective linearity equals 1 for almost perfect nonlinear permutations, equals 2 for random permutations, and is as big as 2^n for affine permutations (where n is the block size). We devise an algorithm that approximately computes the effective linearity of a block cipher with complexity $\Theta(2^{n/2})$. We present various scenarios in which the effective linearity can be used in distinguishing and key recovery attacks. For example, we show that a 2-round Feistel structure surrounded by two key dependent decorrelation modules can be easily distinguished from a random permutation, regardless of the Feistel round function.[4]

This paper is organized as follows: In Section II we give the definition and some theoretical background of highly nonlinear functions. In Section III we present two distinguishing attacks on ciphers with uniform difference distribution tables. Section IV examines differential properties of the construction $g(x) = f^{-1}(f(x) \oplus K)$ and defines the effective linearity of a function. In Section V we use the effective linearity to mount distinguishing and key recovery attacks. In the appendices we present various computations and proofs. We summarize the paper and discuss the implications of our findings in Section VI.

II. Definitions and Theoretical Background

There are several possible notions of high nonlinearity of a boolean function. In this paper we use the following definition, presented in [16]:

Definition 1: A function $f : GF(2^n) \to GF(2^m)$ is perfectly nonlinear if for any non-zero $w \in GF(2^n)$ the difference $f(x+w) - f(x)$ obtains all the values $y \in GF(2^m)$ exactly 2^{n-m} times each.

In [16] the properties of such functions were studied, and it was shown that perfectly nonlinear functions $f : GF(2^n) \to GF(2^m)$ exist if and only if n is even and $n \geq 2m$. However, in real life designs many designers prefer to use functions in which the size of the input is equal to the size of the output. This led to the introduction of *almost perfect nonlinear functions*, defined as following:

Definition 2: A function $f : GF(2^n) \to GF(2^n)$ is almost perfectly nonlinear if for any $a \neq 0$, the function $g(x, a) = f(x) \oplus f(x \oplus a)$ assumes exactly 2^{n-1} different values.

If such f is also a permutation, f is called "almost perfectly nonlinear permutation" (in the sequel we abbreviate this notation to "APNP").

Note, that since the characteristic of the field $GF(2^n)$ is 2, it follows that for any function $f : GF(2^n) \to GF(2^n)$, every value of $g(x, a)$ is assumed an even number of times. Therefore, 2^{n-1} is the maximal possible output size of g for a non-zero α.

A permutation that is close to be an APNP is the S-box SubBytes of AES [8]. This S-box is a permutation $f : GF(2^8) \to GF(2^8)$ and for any $a \neq 0$, the function $g(x, a) = f(x) \oplus f(x \oplus a)$ assumes $2^{8-1} - 1 = 127$ values.

As stated before, this notion of nonlinearity is closely related to differential properties of the function f and the definition of an APNP can be restated in terms of the *difference distribution table* of the function used in differential cryptanalysis [5]. First we recall the definition of the difference distribution table of a function:

Definition 3: Let $f : GF(2^n) \to GF(2^m)$ be a general function. The (XOR) difference distribution table (DDT) of f is an $(2^n) \times (2^m)$ matrix whose (i, j) entry is defined as $\#\{x \in GF(2^n) | f(x) \oplus f(x \oplus i) = j\}$.

A function f is considered optimally secure against differential cryptanalysis if the entries in the DDT of f are the lowest possible ones. This is the case when f is an almost perfect nonlinear function. We can now rephrase the definition of an almost perfect nonlinear function in terms of the DDT:

Definition 4: A function $f : GF(2^n) \to GF(2^n)$ is almost perfectly nonlinear if the highest entry in the DDT of f (except for the entry $(0, 0)$ that equals 2^n) is 2.

This definition is closely related to the following definition of δ-uniformity:

Definition 5: For an $n \times s$ bits S-box $S(\cdot)$ (where $n \geq s$), we denote by δ the highest entry in the difference distribution table (except for the $(0, 0)$ entry), namely

$$\delta = \max_{\alpha \in \{0,1\}^n, \alpha \neq 0, \beta \in \{0,1\}^s} \#\{x | S(x) \oplus S(x \oplus \alpha) = \beta\}.$$

S is called differentially δ-uniform.

Hence, almost perfect nonlinear permutations are differentially 2-uniform.

We recall that differential cryptanalysis is mostly interested in differentials with high probability (or zero probability). This led various papers [18], [19], [23] to suggest using functions that are as differentially uniform as possible. If such functions are used, the cipher is expected to have fewer differentials with high probability, as well as less zero probability differentials.

We emphasize that usually APNPs are not used in real life ciphers. However, this is mostly due to implementation issues,

[3]The notion of "Differential Distribution Table" is defined in Section II.

[4]The notion of "Feistel Structures" is defined in Section II and the notion of "Decorrelation Modules" is defined in Section V-A.

as the common belief is that these functions are better than other constructions. For example, many ciphers use APNPs as building blocks, like the S-box used in the AES.

We conclude this section with a definition of a general block cipher construction that is examined in several places in the sequel:

Definition 6: A Feistel structure $E : GF(2^{2n}) \to GF(2^{2n})$ is a block cipher that can be described in the following manner: The plaintext $P \in GF(2^n)$ is divided into two halves of size n bits each, left (L_0) and right (R_0). Then, the encryption consists of a series of several rounds and in each round the following transformation is applied:

$$L_{i+1} = R_i; R_{i+1} = f_{i+1}(R_i) \oplus L_i,$$

where $f_i(\cdot)$ is the round function of the ith round (the round function may depend on a round subkey). The ciphertext $C = E(P) \in GF(2^{2n})$ is composed of the two halves of the output of the last round.

Finally, we note that throughout the paper we deal with values of n for which 2^{-n} is very close to $1/(2^n - 1)$, and hence we do not distinguish between these values. For small values of n this approximation is inaccurate, but for small block sizes the cipher is susceptible to many attacks, e.g., the dictionary attack, and thus is of little interest.

III. Distinguishing Highly Nonlinear Functions from Random Permutations

In this section we present two distinguishing attacks on highly nonlinear functions. Each is based on a different assumption and performs in a different model (known plaintext or chosen plaintext). These attacks are capable of identifying whether a given black box is a random permutation or an highly nonlinear function. Hence, if an APNP is used as the cipher, it can be distinguished from a random permutation.

One application of our attacks is a distinguishing attack related to the Even-Mansour encryption scheme [9]. This encryption scheme is defined as follows:

Definition 7: Let $P : GF(2^n) \to GF(2^n)$ be a publicly known permutation, and let $K_1, K_2 \in GF(2^n)$ be two secret keys. For every $P \in GF(2^n)$, the encryption of P by the Even-Mansour scheme is $E_{K_1,K_2}(P) = F(P \oplus K_1) \oplus K_2$.

Our distinguisher can be used to distinguish between an Even-Mansour construction using an APNP as the underlying permutation and an Even-Mansour construction using a random permutation as the underlying permutation, despite the commonly believed good security properties of APNPs.

Other possible applications of our attacks are discussed in Section V.

A. A Chosen Plaintext Distinguisher

The first attack is a chosen plaintext attack based on the birthday paradox. The attack uses the following characterization of APNs proposed in [6]:

Proposition 1 ([6], Prop. 2): A function $f : GF(2^n) \to GF(2^n)$ is almost perfectly nonlinear if and only if, for any distinct non-zero $a, b \in GF(2^n)$, we have

$$f(x) \oplus f(x \oplus a) \oplus f(x \oplus b) \oplus f(x \oplus a \oplus b) \neq 0$$

for all $x \in GF(2^n)$.

The distinguisher is summarized in the following proposition:

Proposition 2: Let $f : GF(2^n) \to GF(2^n)$ be a black box permutation for which we have to determine whether it is an APNP or a random permutation. There exists an algorithm with data complexity of $N = 1.4070 \cdot 2^{n/2}$ chosen plaintexts and time complexity of N encryptions and $N/2$ memory accesses, such that:

1) If f is an APNP then the algorithm outputs "APNP" with probability 1.
2) If f is a random permutation, the algorithm outputs "Random Permutation" with probability $p > 0.8$.

In this section we present the distinguishing algorithm, and in the next section we analyze the algorithm and present the proof of Proposition 2.

We perform the following algorithm with a parameter m (to be determined later):

1) Encrypt m distinct pairs of plaintexts (P_1, P_2), such that $P_1 \oplus P_2 = \alpha$ and $P_1 < P_2$ for some fixed non-zero value α and some ordering between the plaintexts $<$ (such as the lexicographical ordering of strings) by f to get the ciphertext pairs of the form $(C_1, C_2) = (f(P_1), f(P_2))$.
2) Store the XOR values of the ciphertexts, i.e., $C_1 \oplus C_2$ in a hash table.
3) If we obtain a collision in the hash table (two pairs with the same ciphertext difference), halt and conclude that f is not an APNP.
4) If no collisions are encountered, conclude that f is an APNP.

A collision is formed of two distinct pairs $(P_1, P_2 = P_1 \oplus \alpha)$ and $(P_3, P_4 = P_3 \oplus \alpha)$, whose corresponding ciphertexts (C_1, C_2) and (C_3, C_4), respectively, satisfy $C_1 \oplus C_2 = C_3 \oplus C_4$. Such a collision means that the equation $f(x) \oplus f(x \oplus \alpha) = C_1 \oplus C_2 = C_3 \oplus C_4$ has (at least) four solutions.

Note that by Proposition 1, if f is an APNP then the equation $f(x) \oplus f(x \oplus \alpha) = \beta$ for non-zero α and β has at most two solutions (x_0 and $x_0 \oplus \alpha$ for some x_0). Thus, for any value of m (even for $m = 2^{n-1}$), no such collision is expected.

For a random permutation, however, the algorithm is expected to find such an instance. Thus, once such an instance is found, the algorithm concludes that f is not an APNP.

B. Analysis of the Chosen Plaintext Attack — Proof of Proposition 2

Let us examine the number of expected collisions for a permutation f. Consider the row corresponding to α in DDT^f, the difference distribution table of f. Recall that $DDT^f(\alpha, \beta)$ (entry (α, β) of DDT^f) is

$$DDT^f(\alpha, \beta) = |\{x \in GF(2^n) : f(x) \oplus f(x \oplus \alpha) = \beta\}|.$$

Hence, for every fixed output difference β, the value that corresponds to β in the considered row represents the number of pairs with input difference α and output difference β, where x and $x \oplus \alpha$ appear as two pairs $(x, x \oplus \alpha)$ and $(x \oplus \alpha, x)$.

For a random permutation f, we may assume that values in any single row of the difference distribution table behave approximately like Poisson random variables [20]. Since the values in each row of the difference distribution table are even and sum up to 2^n, they are distributed according to $2 \cdot Poi(1/2)$. Thus, the value $2k$ is expected to appear in a given row about $2^n \cdot e^{-1/2} \cdot 2^{-k}/k!$ times.

Collisions in Step 3 of the algorithm can occur only for values of β whose corresponding entry of the difference distribution table is greater than 2. Let us examine only values of β in the difference distribution table whose entries are "big", i.e., 4 or more. Out of the 2^n possible β entries, $0.0902 \cdot 2^n$ such entries exist. Let m' be the number of pairs falling into "big" entries, then the probability that no collision is found is

$$\prod_{i=0}^{m'-1}\left(1-\frac{i}{0.0902\cdot 2^n}\right) \leq \prod_{i=0}^{m'-1} e^{-i/0.0902\cdot 2^n}$$

Thus, for a success rate[5] of p, the value of m' should satisfy $p > 1 - e^{-m'\cdot(m'-1)/(2\cdot 0.0902\cdot 2^n)}$. To ensure success probability of 0.8 we need $m' > 0.1618 \cdot 2^{n/2}$ pairs of this kind.

However, the algorithm encrypts also pairs whose output difference β has 2 in the difference distribution table. Hence, the real number of pairs we need to examine is about 4 times larger, as only one out of 4 pairs (more precisely, about 23%) has an output difference meeting our requirement. Therefore, the data complexity of the algorithm is $N = 2m = 1.4070 \cdot 2^{n/2}$ plaintexts (or queries to the black box).

The time complexity of the algorithm is $N = 1.4070 \cdot 2^{n/2}$ encryptions and $m = 0.7035 \cdot 2^{n/2}$ memory accesses. The memory requirements are $m = 0.7035 \cdot 2^{n/2}$ memory cells. This completes the proof of Proposition 2.

The chosen plaintext attack presented above can be converted to a known plaintext attack. In a basic transformation where one α value is considered, $\tilde{m}$ plaintexts are requested. These plaintexts can be combined into about $\tilde{m}^2/2$ pairs, of which a fraction of 2^{-n} have difference α. We proceed with this pairs like in the chosen plaintext attack. Thus, if $\tilde{m} = \sqrt{2 \cdot 2^n \cdot 0.7035 \cdot 2^{n/2}} = 1.1862 \cdot 2^{3n/4}$ then enough pairs with α difference are encountered to ensure a 0.8 success rate.

However, we note that while $\Theta(2^{3n/4})$ known plaintexts ensure that enough plaintexts with a given difference α are encountered, the data set contains enough plaintexts for all values of α. As for the identification of a random permutation it is sufficient that two pairs with some input difference α' and some output difference β' are found, we can mount the attack and examine several differences simultaneously. The data complexity of the resulting attack is $m = 1.3459 \cdot 2^{n/2}$ queries, and the time complexity is $0.9057 \cdot 2^n$. As we present in the improved attack, it is possible to use Wagner's algorithm for the generalized birthday paradox problem [26], reducing the time complexity in exchange for increasing the data complexity.

C. An Improvement of the Chosen Plaintext Attack — Proof of Theorem 1

An improvement of the algorithm uses the fact that the same analysis can be performed for many values of α simultaneously, like in the known plaintext variant of the attack. The improved attack uses m plaintexts, such that the XOR value of any two of them is among a list of m values. For example, setting some of the bits in all the plaintexts to be zero satisfies this property. In this case for each pair of plaintexts (P_1, P_2) we compute $(P_1 \oplus P_2, f(P_1) \oplus f(P_2))$, and insert it into a hash table. A collision in the hash table suggests a quartet of values (P_1, P_2) and (P_3, P_4) such that $P_1 \oplus P_2 = P_3 \oplus P_4$ and $f(P_1) \oplus f(P_2) = f(P_3) \oplus f(P_4)$. This cannot occur for an almost perfect nonlinear permutation, and thus, can be also used for distinguishing.

For m plaintexts chosen in this way, we have $m^2/2$ pairs, each producing a string of $2n$ bits. Not all the 2^{2n} $2n$-bit strings are produced in this process. More precisely, the number of possible values for this string is $m \cdot 2^n$ as there are only m possibilities to the first n bits of the string. If we choose m such that $(m^2/2)^2 > 1.17m2^n$, we have a chance of 50% to find a collision (in case of a random permutation) according to the birthday paradox.[6]

Repeating the previous analysis for a success rate of 0.8, after $m = 1.794 \cdot 2^{n/3}$ pairs the attack is expected to find such a quartet with probability 0.8. Thus, the data complexity of this attack is $m = 1.794 \cdot 2^{n/3}$ chosen plaintexts, and the time complexity of the attack is $m^2/2 = 1.609 \cdot 2^{2n/3}$ memory accesses.

A tradeoff between the data and the time complexities of the attack can be obtained using the algorithm suggested by Wagner in [26] for the generalized birthday problem. Wagner's algorithm addresses the following situation: The attacker is given a list of n-bit words, and has to find four words A, B, C and D such that $A \oplus B \oplus C \oplus D = 0$. The algorithm finds the required quartet in two phases: In the first phase, the attacker finds pairs of words (E, F) in the original list such that the l least significant bits of $E \oplus F$ are equal to 0 (i.e., with the same l least significant bits). In the second phase, the attacker considers the pairs and finds two pairs that have the same value in the remaining bits. Both phases of the algorithm can be implemented efficiently using a hash table. Hence, the algorithm is more efficient once enough quartets exist in the data. As a result, the time complexity of the algorithm is decreased in exchange for an increased data complexity. Note that the basic attack presented above is actually a partial case of Wagner's algorithm, where we fix $l = 0$.

In our chosen plaintext attack, we can apply Wagner's algorithm with words of length $4n/3$, since in all the words we consider there are $2n/3$ bits (out of the $2n$ bits) that are set to zero (in the plaintext). Let $D \geq 1.794 \cdot 2^{n/3}$ be the size of the data set, then the first phase of the algorithm takes about D memory accesses, and outputs a list of about $D^2/2 \cdot 2^{-l}$ pairs which agree on the l least significant bits (taken from the ciphertext side of the string, to assure randomness). Now the

[5]Here and in the sequel "success rate" refers to the probability that if f is a random permutation the distinguisher outputs "Random Permutation".

[6]In general, to achieve a success rate of p the following should be satisfied: $(m^2/2)^2 > \sqrt{2 \cdot \log(1/p)} \cdot m \cdot 2^n$.

attacker searches for a collision in the list of pairs. Using a hash table, this stage takes about $D^2 \cdot 2^{-l-1}$ memory accesses. As there are $4n/3-l$ remaining bits, the probability of finding a collision in the second phase is p if $(D^2/2 \cdot 2^{-l})^2/2 = \sqrt{2\log(1/p)} \cdot 2^{4n/3-l}$. Hence, for a success rate of $p = 0.8$, the data complexity of the attack is $D \approx 1.794 \cdot 2^{n/3+l/4}$. The time complexity of the algorithm is $D + D^2 \cdot 2^{-l-1}$ memory accesses. By choosing $l = 4n/9$, the data complexity is raised to $\Theta(2^{4n/9})$ in exchange for reducing the time complexity to $\Theta(2^{4n/9})$ memory accesses. This completes the proof of Theorem 1.

The situation in the known plaintext attack is similar, but the considered words are of length $2n$ bits. In this case, if the data complexity is $\Theta(2^{n/2+l/4})$, then the time complexity is $\Theta(2^{n/2+l/4} + 2^{n-l/2})$. For example, if $l = 2n/3$, the data complexity and the time complexity are both $\Theta(2^{2n/3})$.

D. Other Kinds of Permutations

Permutations that are very close to APNPs are widely used in block ciphers. For these kinds of permutations, the above attacks still succeed with almost the same success rate. However, in this case even if the function is almost an APNP, the algorithm may yield a wrong answer, though with a very small probability.

Another question which arises, is what happens when the permutation we wish to distinguish is not so close to be differentially 2-uniform. For example, what if there are many entries with value of 4 in the difference distribution table of the permutation. Formally, out of the 2^{n-1} input pairs, let p be an upper bound on the ratio of the pairs in entries with value of 4 in the difference distribution table, while the other non-zero entries are 2 (up to the $0 \to 0$ entry).

For these functions, the above algorithms fail, as the probability to have two pairs whose output difference is the same, is no longer negligible. This problem can be solved when p is far from 0.23,[7] e.g., $p < 0.2$ or $p > 0.3$ (in case $p > 0.3$, we require that at least p of the pairs are in entries with value of 4).

The transformation of the above algorithms to deal with such permutations is changing the identification from "find such an instance, halt and output …" to "count how many instances there are, and compare this number to how many there should be". The analysis of the exact number of plaintexts m needed is quite straightforward given p and the requested success rate. In the next section we shall analyze these scenarios using the effective linearity criterion.

IV. Differential Properties of $f^{-1}(f(x) \oplus K)$ and their Applications

In this section we consider the differential properties of some special structure derived from a permutation f and show how to utilize these properties in order to study the structure of f itself. Let $g(x) = f^{-1}(f(x) \oplus K)$ be a permutation where K is some fixed non-zero key. First, we show that using the properties of g we can determine whether f is an APNP or a random permutation. Then we show how to generalize this result in order to classify functions according to their level of nonlinearity. We formalize this classification by defining the *effective linearity* coefficient (EL) of a permutation that corresponds to the level of linearity determined by our method.

A. Theoretical Background

Let $f : GF(2^n) \to GF(2^n)$ be a black box permutation for which we have to determine whether it is an APNP or a random permutation. Choose an arbitrary non-zero $K \in GF(2^n)$, and define the permutation $g(x) = f^{-1}(f(x) \oplus K)$.

Let A, B be a pair of plaintexts with a non-zero input difference α (e.g., $A \oplus B = \alpha$) and consider $\gamma = g(A) \oplus g(B)$. We shall compute the probability of the event $\gamma = \alpha$, and show that this probability can be used to distinguish APNPs from random permutations.

First, we analyze the case where f is an APNP. We have:

Proposition 3: Let $f : GF(2^n) \to GF(2^n)$ be an APNP and for any pair (A, B) of plaintexts with a non-zero input difference α, let $\gamma = g(A) \oplus g(B)$. Then

$$\Pr_{A,B,K \in GF(2^n), A \neq B, K \neq 0}[\gamma = \alpha] = 2^{-n}. \tag{1}$$

Proof: Let $\beta = f(A) \oplus f(B)$. Since f is an APNP, then (A, B) is the only pair of plaintexts with input difference α and output difference β. Now, consider the pair $(f(A) \oplus K, f(B) \oplus K)$. There are two cases:

1) $f(A) \oplus K = f(B)$. In this case, we have $g(A) = B$ and $g(B) = A$ and thus $\gamma = g(A) \oplus g(B) = B \oplus A = \alpha$. This case occurs when $\beta = f(A) \oplus f(B) = K$ which happens with probability of 2^{-n} (if K is chosen randomly).
2) $f(A) \oplus K \neq f(B)$. In this case, the pairs $(f(A), f(B))$ and $(f(A) \oplus K, f(B) \oplus K)$ differ, but still have the same XOR difference β. Thus, since f is an APNP, the difference $\gamma = g(A) \oplus g(B) = (f^{-1}(f(A) \oplus K)) \oplus (f^{-1}(f(B) \oplus K))$ cannot be equal to α. Therefore, in this case $\gamma \neq \alpha$ always.

Combining the two cases together we obtain

$$\Pr_{A,B,K \in GF(2^n), A \neq B, K \neq 0}[\gamma = \alpha] = 2^{-n},$$

as asserted. ■

The analysis presented in Appendix I shows that for a random permutation this probability ($\Pr[\gamma = \alpha]$) equals to $2 \cdot 2^{-n}$. The difference between the probabilities can be used in order to distinguish between an APNP and a random permutation.

We have experimentally verified that the value $2 \cdot 2^{-n}$ is the correct value for a random permutation and that 2^{-n} is the correct value for an APNP. This was done by generating sets of random permutations of 8,10,12,14 and 16 bits, and counting all possible quartets $(A, B, g(A), g(B))$ (for a large set of K values, for all α values).

[7] The expected ratio of 4 or more in the difference distribution table of a random permutation is about 0.09 of the entries. However, entries with 6,8, or even more, contribute more collisions. The total number of collisions counted by the above algorithms for a random permutation is equal to the case where 0.23 of the entries of the table are 4's (while the remaining are 2's and 0's).

B. An Adaptive Chosen Plaintext and Ciphertext Distinguisher

The algorithm of the distinguisher is as follows: Let $f : GF(2^n) \to GF(2^n)$ be a black box permutation for which we have to determine whether it is an APNP or a random permutation and let $m, threshold$ be integers specified later.

1) Ask for the encryption of m distinct plaintexts P_i, for $i = 1, \ldots, m$.
2) Choose $K \in GF(2^n)$ randomly and uniformly.
3) Ask for the decryption of the values $f(P_i) \oplus K$ and obtain $g(P_i) = f^{-1}(f(P_i) \oplus K)$.
4) Store the m values of the form $P_i \oplus g(P_i)$ into a hash table.
5) Count the number of collisions in the hash table. If the number of collisions is greater than $threshold$ output "random permutation". Otherwise, output "APNP".

We note that if $P_1 \oplus g(P_1) = P_2 \oplus g(P_2)$ then the quartet $((P_1, P_2), (g(P_1), g(P_2)))$ satisfies $\gamma = \alpha$. In the sequel we call such quartets "right quartets". Starting with m plaintexts, we get m values of $P \oplus g(P)$. Once there is a collision in the hash table, the colliding values suggest a quartet. If there are three values in the same entry of the hash table, then we get three quartets, or generally, if there are k values in the same entry, we get $k(k-1)/2$ quartets.

For an APNP, a fraction of 2^{-n} of all the possible quartets satisfy our conditions, while for a random permutation, a fraction of $2 \cdot 2^{-n}$ satisfy our conditions. For $m = 4 \cdot 2^{n/2}$, the expected number of quartets for a random permutation is 16, while it is only 8 for an APNP. As the number of expected quartets has a Poisson distribution, for $threshold = 10$ the algorithm outputs the correct answer with probability $p > 0.816$. Similar analysis can be applied for $m = 2^{n/2}$, for which the correct answer is output with probability $p' > 0.594$.

This attack may seem less desirable, given the attacks of the previous section, as it has a similar or even larger data complexity but a more stern attack model. However, this attack can be easily extended to other cases, as we present in Section IV-C.

We remark that one can make a slight change in the attack such that only quartets of the form $(A, B, g(A), g(B))$ where $A \neq B \neq g(A) \neq g(B)$ are counted. Using this variant of the attack the number of expected collisions is 0 for APNP, and if we get even one collision the permutation is certainly not an APNP. The number of expected collisions for a random permutation is a fraction of $\frac{1}{4} \cdot 2^{-n}$ of the total number of possible quartets.

This can be used to increase the success probability of the attack by setting $threshold = 1$. More accurately, for the same data complexity as before, the success rate of the attack is about 0.98. We can also reduce the data complexity by a factor of $\sqrt{2}$ and still have a success rate of 0.86.

C. The Effective Linearity of a Permutation

Following the previous attack, we define the effective linearity of a permutation.

Definition 8: Let $f : GF(2^n) \to GF(2^n)$ be a permutation. The effective linearity of f is:

$$EL(f) = 2^n \left(\frac{1}{2^n - 1}\right)^2 \cdot \sum_{\alpha, K \in GF(2^n) \setminus \{0\}} \Pr\left[\alpha \xrightarrow{g(x) = f^{-1}(f(x) \oplus K)} \alpha\right]$$

Actually, $EL(f)$ is the average of the probabilities $\Pr[\alpha = \gamma]$ over all non-zero K's and α's scaled by a factor of 2^n. It appears that there is a close relation between the effective linearity of f and its difference distribution table DDT^f, as illustrated by the following theorem, whose proof appears in Appendix I:

Theorem 2: Let $f : GF(2^n) \to GF(2^n)$ be a permutation. Then

$$EL(f) \approx -1 + 2^{-2n} \cdot \sum_{\alpha, \beta \in GF(2^n) \setminus \{0\}} [DDT^f(\alpha, \beta)]^2. \quad (2)$$

Hence, the algorithm presented in Section IV-B allows to evaluate a general measure on the differential structure of f with a relatively small complexity of $\Theta(2^{n/2})$, where n is the plaintext size.

For a random permutation f the value $EL(f)$ is expected to be close to 2 (as shown in Appendix I). Hence, if the value of EL of some permutation is not close to 2, then our attack can be applied to the permutation and distinguish it from a random permutation.

We can either calculate $EL(f)$ analytically when the difference distribution table of f is known (like in the analysis for a random permutation presented in Appendix I), or by experimentally measuring it. Taking several sets of $\Theta(2^{n/2})$ messages and using several K values, we can use statistical methods to evaluate $EL(f)$. Note that the $\Theta(2^{n/2})$ complexity is achieved by using many (if not all) values of α simultaneously.

The effective linearity of f is not lower than 1 (as when $K = f(A) \oplus f(B)$ we get that $g(A) = B$ and $g(B) = A$), and cannot be higher than 2^n (which is the value for affine permutations). As the value for a random permutation is 2, we suggest designing ciphers with effective linearity close to 2.

It is possible to show (see Appendix II) that for a two round Feistel construction whose round functions are both APNPs, the effective linearity is 7, while if the round functions are random permutations it is expected to be 14.

An interesting remark about Feistel constructions is that if the permutation of the first round p_1 has effective linearity $EL(p_1)$, and the second round's permutation p_2 has effective linearity $EL(p_2)$, then the effective linearity $EL(f)$ of the two round Feistel construction satisfies $EL(f) \geq EL(p_1) \cdot EL(p_2)$. The proof is given in Appendix II.

When the round functions are not bijective, the difference distribution table of the 2-round Feistel construction is expected to have more zero entries than usual. As the sum of every line in the difference distribution table is constant, it follows that the remaining entries are expected to be higher, leading to an higher effective linearity. Thus, the more entries having a zero value in the difference distribution table, the higher the effective linearity is expected to be.

V. Various Attacks Based on the Effective Linearity of Permutations

In this section we present several possible scenarios in which measuring the effective linearity of various permutations can be used in order to mount distinguishing and key recovery attacks.

A. Treating Decorrelation Modules

The first application of our distinguishing attacks is treating encryption schemes based on decorrelation modules. Decorrelation modules, introduced by Vaudenay [24], are defined as follows:

Definition 9: A decorrelation module of order d is a keyed permutation which ensures that any statistical property of order d (i.e., that uses at most d input/output pairs) averaged over all the keys becomes similar to the random case. For example, the following permutation $DM : GF(2^n) \to GF(2^n)$ is a decorrelation module of order two: $DM_{K_1,K_2}(X) = (X \oplus K_1) \cdot K_2$, where the multiplication is done over the field $GF(2^n)$, and $K_2 \neq 0$. It means that any adversary that can obtain two input/output pairs of this decorrelation module, cannot distinguish (in the information theoretic sense, i.e., independent of its computational power) whether they originated from the module or from a random permutation.

One property of decorrelation modules is that once the key is set the decorrelation module is linear, but when the key is random, the probability of any non-trivial differential going through the module equals $1/(2^n - 1)$ on average. A similar condition can be proved with respect to linear cryptanalysis as well.

Consider a cipher of the form $E = DM_2 \circ F \circ DM_1$, where DM_i $(i = 1, 2)$ is a keyed decorrelation module and F is a (possibly weak) non-linear permutation.

Due to the nature of decorrelation modules, any differential (even a truncated one) cannot have probability higher than the trivial one through the first decorrelation module. The same holds for linear approximations as well. Hence, the first decorrelation module makes the cipher immune to differential and linear attacks in the forward direction. Similarly, the second decorrelation module makes the cipher immune to differential and linear attacks in the backward direction. Other advanced attacks also cannot be applied, even if F is cryptographically weak, since the attacker cannot control the access to F due to the decorrelation modules.

While all these methods fail, our technique can use the weakness of F to mount a distinguishing attack on the entire construction. Indeed, since for any fixed key decorrelation modules are linear, and since the effective linearity computes the average over all the possible differentials, application of a decorrelation module as the first or the last operation in the encryption process does not affect the effective linearity. Hence, the effective linearity of the entire construction is equal to the effective linearity of F, and therefore if F is relatively weak, this can be used to distinguish between the entire construction and a random permutation.

For example, consider a cipher of the form $E = DM_2 \circ F_2 \circ F_1 \circ DM_1$, where DM_i $(i = 1, 2)$ is a decorrelation module, and F_i is a Feistel round with a random permutation as the round function (along with some key).

While all the standard methods fail to attack this construction, we can efficiently distinguish the above E from a random permutation. As explained above, $EL(E)$ is equal to the effective linearity of the inner two-round Feistel construction, and since the round functions of the Feistel construction are random permutations, we get $EL(E) = 14$. As the effective linearity of a random permutation is only 2, we can distinguish between E and a random permutation with data and time complexity of $\Theta(2^{n/2})$ queries.

We note that the minimal value of the effective linearity of a 2-round Feistel construction is 7 (achieved by applying APNPs as the round functions in the two Feistel rounds). Thus, even if the Feistel round functions are replaced, our attack still succeeds.

B. Distinguishing Known Ciphers and Identifying Black Box Permutations

It is possible to precompute the effective linearity of several known ciphers in advance (also for reduced round variants). Then, given a black box the attacker computes its EL and if the black box is one of the previously known encryption schemes, he can detect it.

However, we still do not know whether this attack is applicable against encryption schemes that are actually used today.[8] It may occur that the measured effective linearity values are too close to 2 and then the distinguishing is infeasible. Moreover, as we note in Appendix I, the number of quartets detected in the distinguisher slightly depends on K and thus in some cases the difference between two encryption schemes can be smaller than the difference between applications of the same scheme with different values of K. In this case the attack might fail.

The effective linearity of a permutation depends on its difference distribution table. When computing the difference distribution table of a permutation one usually computes the average probabilities over all the possible keys and assumes that the probability for any single key is close to the average (see [5]). However, in some ciphers there are classes of keys for which some differential properties of the cipher differ from the average case, like for IDEA [7]. Such classes are called *differentially weak key classes*. Usually such classes can be detected only if some explicit differential characteristic is known for the whole cipher.

We can use our distinguisher in order to detect such classes when the entire differential structure differs for different keys, even if no concrete characteristic is known. For example, there is a differential weak key class of IDEA. In that weak key class there exists some differential with probability 1. For this weak key class, the effective linearity is higher by 1 than the effective linearity of IDEA with a key that does not belong to this weak key class. We note that using the differential is easier for the purpose of distinguishing whether the key is in the

[8] We note that evaluating the EL analytically is usually infeasible for more than a few rounds of the cipher, and approximating it using the algorithm presented in Section IV-B is feasible only for ciphers with blocks of length at most 64 bits.

weak key class. However, if decorrelation modules are added before and after IDEA, then the differential distinguisher is no longer applicable, while our distinguisher still succeeds.

Our technique can be also used in key recovery attacks, by measuring the effective linearity of reduced round variants of the cipher. Then, for a given r-round construction, we can guess the subkey of the last round, and try to decrypt the ciphertexts through the last round. If the subkey guess is correct, the effective linearity of the obtained cipher equals to the one of $(r-1)$ rounds of the original cipher, instead of the expected $(r+1)$ rounds in case of a wrong guess.

C. Attacks Against Encryption Schemes of the Form $f^{-1}(f(x) \oplus K)$

The distinguisher described in Section IV-B can be applied directly to distinguish encryption schemes of the form $h(x) = f^{-1}(f(x) \oplus K)$ (for an arbitrary permutation f, when K is a secret key) from a random permutation. In this scenario, the data requirements are even less restrictive than in the original distinguisher: the attacker can use known plaintexts instead of adaptively chosen plaintexts and ciphertexts.

The algorithm of the distinguisher in this case is the following (where m and $threshold$ are specified later):

1) Encrypt m distinct plaintexts $\{P_i\}_{i=1}^{m}$ through h to get the values $\{h(P_i)\}_{i=1}^{m}$.
2) Store the m values of the form $P_i \oplus h(P_i)$ in a hash table.
3) Count the number of collisions in the hash table. If the number of collisions is greater than $threshold$ output "h of the form $h(x) = f^{-1}(f(x) \oplus K)$". Otherwise, output "random permutation".

Recall that collisions in the hash table correspond to pairs (P_1, P_2) such that $P_1 \oplus P_2 = h(P_1) \oplus h(P_2)$. Hence, if h is a random permutation then the expected fraction of collisions amongst the possible pairs is 2^{-n}.

On the other hand, if h is of the form $h(x) = f^{-1}(f(x) \oplus K)$, then as shown in Section IV-B, the number of collisions is related to the effective linearity of f. In particular, if f is a random permutation, the expected fraction of collisions amongst the possible pairs is $2 \cdot 2^{-n}$.

Therefore, if we start with $m = 4 \cdot 2^{n/2}$ plaintexts, the expected number of collisions for h of the form $h(x) = f^{-1}(f(x) \oplus K)$ is 16, while for a random permutation the expected number is only 8. As the number of expected quartets has a Poisson distribution, for $threshold = 10$ the algorithm outputs the correct answer with probability $p > 0.816$. Similar analysis can be applied for $m = 2^{n/2}$, for which the correct answer is output with probability $p' > 0.594$.

We remark that constructions of the form $h(x) = f^{-1}(f(x) \oplus K)$ are not so rare. For example, two rounds of any involution cipher, such as KHAZAD [2] are of this form.

Another example of a construction of the form $h(x) = f^{-1}(f(x) \oplus K)$ is any permutation $P : GF(2^n) \rightarrow GF(2^n)$ such that for every $x \in GF(2^n)$ there exists $y \neq x$ such that $P(x) = y$ and $P(y) = x$. Indeed, for such permutation we can divide the elements of $GF(2^n)$ to pairs (x, y) such that $P(x) = y$ and $P(y) = x$, and arrange the pairs according to some ordering, both between the pairs and inside each pair (e.g., the lexicographic ordering is good for this purpose). After the ordering, each element $x \in GF(2^n)$ is indexed by an n-bit vector, where the first $n-1$ bits represent the number of the pair x belongs to, and the last bit represents the position of x in its pair. Then we can define $f : GF(2^n) \rightarrow GF(2^n)$ (arranged by the ordering) by $f(x_1, \ldots, x_n) = (x_1, \ldots, x_{n-1}, 1 - x_n)$. It is easy to see that for every $x \in GF(2^n)$ we have $P(x) = f^{-1}(f(x) \oplus (0, 0, \ldots, 0, 1))$. Hence, our attack is applicable to any cipher of this structure.

Construction	Round Function	Effective Linearty
APNP	APNP	1
Random Permutation	Random Permutation	2
Affine Permutation	Affine Permutation	2^n
DES	1 Round	$2^{40.32}$
DES	2 Rounds	$2^{16.65}$
AES	1 Round	$2^{25.54}$
AES	2 Rounds	≥ 49 (proved) 244 (conjectured)
2-round Feistel	APNP	7
	Random Permutation	14

TABLE I
VARIOUS CONSTRUCTIONS AND THEIR EFFECTIVE LINEARITIES

VI. SUMMARY

In the first part of this paper we presented several distinguishers between highly nonlinear permutations and random permutations. We conclude that while using APNPs as part of the encryption scheme seems desirable, using APNPs as the entire cipher can possess inherent weakness.

In the second part of the paper we examined the construction $g(x) = f^{-1}(f(x) \oplus K)$ for various permutations f. We showed how to use the differential properties of this construction in order to study the differential structure of f. We also proved that this construction can be used to effectively determine the average of the squares of the probabilities of differentials through f. Finally, we defined the *effective linearity* of a permutation that measures this probability. Table I contains various constructions and their effective linearity.

The effective linearity can be used to distinguish between an f that is an almost perfect nonlinear permutation and an f that is a random permutation. It can also be used to distinguish ciphers with a relatively close to linear structure from random permutations, even if no concrete statistical property is known. For several constructions, our attacks have better performance compared to previously known attacks. For example, we can distinguish between a random permutation and a permutation formed by two Feistel rounds surrounded by two key-dependent decorrelation modules, regardless of the round functions of the 2-round Feistel construction.

REFERENCES

[1] Kazumaro Aoki, Serge Vaudenay, *On the Use of GF-Inversion as a Cryptographic Primitive*, proceedings of Selected Areas in Cryptography 2003, Lecture Notes in Computer Science 3006, pp. 234–247, Springer-Verlag, 2004.

[2] Paulo S.L.M. Baretto, Vincent Rijmen, *The KHAZAD Block Cipher*, Submitted to NESSIE, available online at *http://www.nessie.eu.org*.
[3] Thomas Beth, Cunsheng Ding, *On Almost Perfect Nonlinear Permutations*, Advances in Cryptography, proceedings of EUROCRYPT '93, Lecture Notes in Computer Science 765, pp. 65–76, Springer-Verlag, 1994.
[4] Eli Biham, Alex Biryukov, Adi Shamir, *Cryptanalysis of Skipjack reduced to 31 rounds*, Advances in Cryptology, proceedings of EUROCRYPT '99, Lecture Notes in Computer Science 1592, pp. 12–23, Springer-Verlag, 1999.
[5] Eli Biham, Adi Shamir, *Differential Cryptanalysis of the Data Encryption Standard*, Springer-Verlag, 1993.
[6] Anne Canteaut, *Cryptographic Functions and Design Criteria for Block Ciphers*, proceedings of Indocrypt 2001, Lecture Notes in Computer Science 2247, pp. 1–16, Springer-Verlag, 2001.
[7] Joan Daemen, Rene Govaerts, Joos Vandewalle, *Weak Keys for IDEA*, Advances in Cryptology, proceedings of CRYPTO '93, Lecture Notes in Computer Science 773, pp. 224–231, Springer-Verlag, 1994.
[8] Joan Daemen, Vincent Rijmen, *The design of Rijndael: AES — the Advanced Encryption Standard*, Springer-Verlag, 2002.
[9] Shimon Even, Yishay Mansour, *A Construction of a Cipher from a Single Pseudorandom Permutation*, Journal of Cryptology, Vol. 10, Number 4, pp. 151–162, Springer-Verlag, 1997.
[10] Philip Hawkes, Gregory G. Rose, *Primitive Specification for SOBER-t16 Submission to NESSIE* and *Primitive Specification for SOBER-t32 Submission to NESSIE*, Submitted to NESSIE, available online at *http://www.nessie.eu.org*.
[11] Michael Luby, Charles Rackoff, *How to Construct Pseudorandom Permutations from Pseudorandom Functions*, SIAM journal of Computing, Volume 17. Number 2, pp. 373–386, 1988.
[12] Mitsuru Matsui, *Linear Cryptanalysis Method for DES Cipher*, Advances in Cryptology, proceedings of EUROCRYPT '93, Lecture Notes in Computer Science 765, pp. 386–397, Springer-Verlag, 1994.
[13] Willi Meier, Othmar Staffelbach, *Nonlinearity Criteria for Cryptographic Functions*, Advances in Cryptology, proceedings of EUROCRYPT '89, Lecture Notes in Computer Science 434, pp. 549–562, Springer-Verlag, 1990.
[14] Willi Meier, Othmar Staffelbach, *Fast Correlation Attacks on Stream Ciphers (Extended Abstract)*, Advances in Cryptology, proceedings of EUROCRYPT '88, Lecture Notes in Computer Science 330, pp. 300–315, Springer-Verlag, 1988.
[15] US National Bureau of Standards, *Data Encryption Standard*, Federal Information Processing Standards Publications Number 46, 1977.
[16] Kaisa Nyberg, *Perfect nonlinear S-boxes*, Advances in Cryptology, proceedings of EUROCRYPT '91, Lecture Notes in Computer Science 547, pp. 378–386, Springer-Verlag, 1991.
[17] Kaisa Nyberg, *On the construction of highly nonlinear permutations*, Advances in Cryptology, proceedings of EUROCRYPT '92, Lecture Notes in Computer Science 658, pp. 92–98, Springer-Verlag, 1993.
[18] Kaisa Nyberg, *Differentially uniform mappings for cryptography*, Advances in Cryptology, proceedings of EUROCRYPT '93, Lecture Notes in Computer Science 765, pp. 55–64, Springer-Verlag, 1994.
[19] Kaisa Nyberg, Lars R. Knudsen, *Provable Security Against Differential Cryptanalysis*, Advances in Cryptology, proceedings of CRYPTO '92, Lecture Notes in Computer Science 740, pp. 566–578, Springer-Verlag, 1993.
[20] Luke O'Connor, *On the Distribution of Characteristics in Bijective Mappings*, Advances in Cryptology, proceedings of EUROCRYPT'93, Lecture Notes in Computer Science 765, pp. 360–370, Springer-Verlag, 1994.
[21] Oscar S. Rothaus, *On Bent Functions*, Journal of Combinatorial Theory, Series A, Vol. 20 (1976), pp. 305–310, 1976.
[22] Jennifer Seberry, Xian-Mo Zhang, Yuliang Zheng, *Relationships Among Nonlinearity Criteria (Extended Abstract)*, Advances in Cryptology, proceedings of EUROCRYPT '94, Lecture Notes in Computer Science 950, pp. 376–388, Springer-Verlag, 1995.
[23] Jennifer Seberry, Xian-Mo Zhang, Yuliang Zheng, *Pitfalls in Designing Substitution Boxes (Extended Abstract)*, Advances in Cryptology, proceedings of CRYPTO '94, Lecture Notes in Computer Science 839, pp. 383–396, Springer-Verlag, 1995.
[24] Serge Vaudenay, *Provable Security for Block Ciphers by Decorrelation*, Journal of Cryptology, Vol. 16, Number 4, pp. 249–286, Springer-Verlag, 2003.
[25] David Wagner, *The Boomerang Attack*, proceedings of Fast Software Encryption 6, Lecture Notes in Computer Science 1636, pp. 156–170, Springer-Verlag, 1999.
[26] David Wagner, *A Generalized Birthday Problem (Extended Abstract)*, Advances in Cryptology, proceedings of CRYPTO '02, Lecture Notes in Computer Science 2442, pp. 288–304, Springer-Verlag, 2002.

APPENDIX I
THEORETICAL ANALYSIS OF THE EFFECTIVE LINEARITY OF VARIOUS PERMUTATIONS

Let f be a permutation and let $g(x) = f^{-1}(f(x) \oplus K)$ for some randomly chosen K. In this appendix we analyze the number of quartets $(A, B, g(A), g(B))$ satisfying $A \oplus B = g(A) \oplus g(B)$. We show that the number of such quartets (called in the sequel "right quartets") is related to the effective linearity of f, as well as to the difference distribution table of f. First, we present the proof of Theorem 2 and then we apply the theorem to compute the effective linearity of various constructions.

Examine the amount of quartets of the form $(A, B, g(A), g(B))$ such that $\gamma \triangleq g(A) \oplus g(B) = A \oplus B = \alpha$. We have two cases:

When $g(A) = B$ then $g(B) = A$ as well:

$$g(A) = f^{-1}(f(A) \oplus K) = B \Rightarrow f(B) = f(A) \oplus K \Rightarrow$$

$$f(A) = f(B) \oplus K \Rightarrow A = f^{-1}(f(B) \oplus K) = g(B).$$

Hence, in this case $\gamma = \alpha$ holds with probability 1. As K is chosen randomly, there is probability 2^{-n} that $f(A) \oplus f(B) = K$, and therefore, this case occurs with probability 2^{-n}.

When $g(A) \neq B$ all the elements of the quartet $(A, B, g(A), g(B))$ are distinct. We observe that any quartet of this kind can be uniquely represented by the two ciphertext pairs $(f(A), f(B))$, $(f(A) \oplus K, f(B) \oplus K)$. These two pairs satisfy the following system of equations:

$$\begin{cases} x \oplus y = \beta \\ f^{-1}(x) \oplus f^{-1}(y) = \alpha \end{cases} \tag{3}$$

for some fixed α and β.

On the other hand, consider pairs of inputs to the function f^{-1} that solve System (3) for some fixed α and β. Every pair of such pairs (x_1, y_1), (x_2, y_2) represents a right quartet if and only if one of the following holds:

$$\begin{array}{c} x_1 \oplus x_2 = y_1 \oplus y_2 = K \\ \text{or} \\ x_1 \oplus y_2 = y_1 \oplus x_2 = K \end{array} \tag{4}$$

The probability of this event for the pair of pairs is about $2 \cdot 2^{-n}$.

For sake of simplicity we assume that this probability is independent of K, but actually, there is a measure of dependence on K. The same computation can be rewritten in another way such that the dependence on α is neglected and the dependence on K is computed explicitly. The resulting formula is similar to Equation (7) below when K is substituted instead of α. This is the case in the attacks presented in Section IV-B since we look for the average of the results for different α values, but for the same value of K. Since the expected result considers the average for all possible keys, changing the formulae to be dependent on K does not affect the expected results.

In order to compute the probability of a quartet to be a right quartet, we go over all the possible quartets $(A, B, g(A), g(B))$ such that $A \oplus B = \alpha$ and count the number of right quartets. Note that since every value of A determines uniquely a quartet, and it follows that the total number of quartets is 2^n.

Denote by t_α the number of un-ordered pairs of solutions of System (3) for the specific value α, summed over all possible β's. Then the expected number of pairs of solutions that satisfy one of the conditions of Equation (4) is $2 \cdot 2^{-n} \cdot t_\alpha$. However, each such un-ordered pair of solutions $\{(x_1, y_1), (x_2, y_2)\}$ leads to four right quartets:

1) $((f^{-1}(x_1), f^{-1}(y_1)), (f^{-1}(x_2), f^{-1}(y_2)))$,
2) $((f^{-1}(x_2), f^{-1}(y_2)), (f^{-1}(x_1), f^{-1}(y_1)))$,
3) $((f^{-1}(y_1), f^{-1}(x_1)), (f^{-1}(y_2), f^{-1}(x_2)))$,
4) $((f^{-1}(y_2), f^{-1}(x_2)), (f^{-1}(y_1), f^{-1}(x_1)))$.

Hence, the expected number of right quartets encountered while trying all possible quartets is $8 \cdot 2^{-n} \cdot t_\alpha$. Finally, since the total number of quartets is 2^n, the probability of a single quartet to be a right quartet is $8 \cdot 2^{-2n} \cdot t_\alpha$.

In order to compute the value of t_α, we consider the function f^{-1}. For some fixed value β_0 and the element of the difference distribution table of f^{-1} corresponding to the pair (β_0, α), let $DDT^{f^{-1}}(\beta_0, \alpha) = 2k$. In this case, there are k solutions of the system (3) and thus there are $k(k-1)/2$ distinct un-ordered pairs of solutions. Summing over all the possible values of β, we obtain that

$$t_\alpha = \sum_{\beta \in GF(2^n)} \frac{DDT^{f^{-1}}(\beta,\alpha)/2 \cdot (DDT^{f^{-1}}(\beta,\alpha)/2 - 1)}{2}, \qquad (5)$$

where $DDT^{f^{-1}}$ is the difference distribution table of f^{-1}.

Recall that the difference distribution table of the function f^{-1} is actually the transpose of the difference distribution table of f, i.e., $(DDT^f)^T = DDT^{f^{-1}}$. Thus, we are able to rewrite Equation (5) in terms of the difference distribution table of f:

$$\begin{aligned} t_\alpha &= \sum_{\beta \in GF(2^n)} \frac{DDT^f(\alpha,\beta)/2 \cdot (DDT^f(\alpha,\beta)/2 - 1)}{2} = \qquad (6) \\ &\frac{1}{8} \cdot \sum_{\beta \in GF(2^n)} DDT^f(\alpha,\beta) \cdot \left(DDT^f(\alpha,\beta) - 2\right) = \\ &\frac{1}{8} \cdot \sum_{\beta \in GF(2^n)} \left[DDT^f(\alpha,\beta)\right]^2 - 2 \cdot DDT^f(\alpha,\beta) = \\ &-\frac{1}{4} \cdot 2^n + \frac{1}{8} \cdot \sum_{\beta \in GF(2^n)} \left[DDT^f(\alpha,\beta)\right]^2 \end{aligned}$$

We conclude that:

$$\begin{aligned} EL(f) &= 2^n \left(\frac{1}{2^n - 1}\right)^2 \cdot \sum_{\alpha, K \in GF(2^n)\setminus\{0\}} \Pr\left[\alpha \xrightarrow{g(x)=f^{-1}(f(x)\oplus K)} \alpha\right] = \qquad (7) \\ &2^n \left(\frac{1}{2^n - 1}\right)^2 \cdot \sum_{\alpha, K \in GF(2^n)\setminus\{0\}} 2^{-n} + 8 \cdot 2^{-2n} \cdot t_\alpha = \\ &\left(\frac{1}{2^n - 1}\right)^2 \cdot \sum_{\alpha, K \in GF(2^n)\setminus\{0\}} 1 + 8 \cdot 2^{-n} \cdot t_\alpha \approx 1 + \frac{8\overline{t_\alpha}}{2^n} \approx \\ &-1 + 2^{-2n} \cdot \sum_{\alpha,\beta \in GF(2^n), \alpha \neq 0} [DDT^f(\alpha,\beta)]^2 \end{aligned}$$

where $\overline{t_\alpha}$ is the average value of t_α taken over all α values.

This completes the proof of Theorem 2.

We note that a differential $\alpha \to \beta$ with probability p contributes p^2 to the effective linearity. Indeed, the entry (α, β) in the difference distribution table has a value of $p \cdot 2^n$, and thus, this entry contributes $2^{-2n} \cdot (p \cdot 2^n)^2 = 2^{-2n} \cdot p^2 \cdot 2^{2n} = p^2$ to the effective linearity.

A. *Effective Linearity of Affine Permutations*

In this section we compute the effective linearity of affine permutations.

Proposition 4: Let f be an affine permutation over $GF(2^n)$, i.e., $f(x) = A \cdot x \oplus b$ for some fixed constants $A \in GF(2^{n \times n}), b \in GF(2^n)$. Then $EL(f) = 2^n$.

Proof: The difference distribution table of an affine permutation is very simple. For every $\Delta \in GF(2^n)$, we have

$$f(x \oplus \Delta) \oplus f(x) = [A \cdot (x \oplus \Delta) \oplus b] \oplus [A \cdot x \oplus b] = A \cdot [(x \oplus \Delta) \oplus x] = A \cdot \Delta.$$

This means, that given an input difference Δ, the output difference is always $A \cdot \Delta$. As A is invertible (since otherwise f is not a permutation), multiplication by A is a permutation, i.e., the output difference $A \cdot \Delta$ can be reached only by the input difference Δ. Therefore, each row and each column in the difference distribution table of f has one entry with the value 2^n, while all the rest are 0's.

When considering the permutation $g(x) = f^{-1}(f(x) \oplus K)$, for any value of K the following holds:

$$g(x \oplus \alpha) \oplus g(x) = f^{-1}(f(x \oplus \alpha) \oplus K) \oplus f^{-1}(f(x) \oplus K).$$

If $f(x) = A \cdot x \oplus b$, then $f^{-1}(x) = A^{-1} \cdot (x \oplus b) = A^{-1} \cdot x \oplus A^{-1} \cdot b$. Hence,

$$g(x \oplus \alpha) \oplus g(x) = f^{-1}\left(f(x) \oplus K \oplus A \cdot \alpha\right) \oplus f^{-1}\left(f(x) \oplus K\right) =$$

$$f^{-1}(b \oplus A \cdot \alpha) \oplus f^{-1}(b) = A^{-1} \cdot A \cdot \alpha = \alpha$$

Therefore, the effective linearity of f is:

$$EL(f) = 2^n \left(\frac{1}{2^n - 1}\right)^2 \cdot \sum_{\alpha, K \in GF(2^n)\setminus\{0\}} \Pr\left[\alpha \xrightarrow{g(x)=f^{-1}(f(x)\oplus K)} \alpha\right] =$$

$$2^n \left(\frac{1}{2^n - 1}\right)^2 \cdot \sum_{\alpha, K \in GF(2^n)\setminus\{0\}} 1 = 2^n,$$

as asserted. ■

B. *Effective Linearity of Almost Perfect Nonlinear Permutations*

The difference distribution table of an APNP contains only 2's and 0's (besides the $0 \to 0$ entry). Hence, using Equation (6), we get that for almost perfect nonlinear permutations, for any non-zero α:

$$t_\alpha = \sum_{\beta \in GF(2^n)} \frac{DDT^f(\alpha,\beta)/2 \cdot (DDT^f(\alpha,\beta)/2 - 1)}{2} =$$

$$2^{n-1} \cdot \frac{(2/2) \cdot (2/2 - 1)}{2} = 0$$

Thus, the only case for which a quartet $A, B, g(A), g(B)$ satisfies $A \oplus B = g(A) \oplus g(B)$ is when $A = g(B)$. This happens with probability 2^{-n}, and thus, $EL(APNP) = 1$.

C. *Effective Linearity of Random Permutations*

In this section we compute the effective linearity of a random permutation, i.e., a permutation f chosen at random from all the permutations of $GF(2^n)$.

Proposition 5: Let $f : GF(2^n) \to GF(2^n)$ be a random permutation. Then EL(f)=2.

Proof: We may assume that values in any single row[9] of the difference distribution table of f are distributed approximately as Poisson random variables. That is, the values in the difference distribution table are distributed according to $2 \cdot Poi(1/2)$.[10] Thus, the value $2k$ is expected to appear in a given row about $2^n \cdot e^{-1/2} \cdot 2^{-k}/k!$ times. Substituting these figures to Equation (6), we obtain that for a random permutation:

$$t_\alpha = \sum_{k=1}^{2^{n-1}} \frac{2^n \cdot e^{-1/2} \cdot 2^{-k}}{k!} \cdot \frac{k \cdot (k-1)}{2} = 1/2 \cdot 2^n \left[\underbrace{\sum_{k=1}^{2^{n-1}} \frac{k^2 e^{-1/2} 2^{-k}}{k!}}_{I_1} - \underbrace{\sum_{k=1}^{2^{n-1}} \frac{k e^{-1/2} 2^{-k}}{k!}}_{I_2} \right] \tag{8}$$

Let X be a random variable distributed according to $Poi(\frac{1}{2})$, then we have $I_1 = E[X^2]$ and $I_2 = E[X]$. For such X, $E[X] = \frac{1}{2}$ and $Var[X] = E[X^2] - E[X]^2 = \frac{1}{2}$. Thus, $E[X^2] = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$. Hence, we get

$$t_\alpha = 1/2 \cdot 2^n \left[\sum_{k=1}^{2^{n-1}} \frac{k^2 e^{-1/2} 2^{-k}}{k!} - \sum_{k=1}^{2^{n-1}} \frac{k e^{-1/2} 2^{-k}}{k!} \right] = 1/2 \cdot 2^n \left[\frac{3}{4} - \frac{1}{2} \right] = 1/8 \cdot 2^n \tag{9}$$

Therefore, according to Equation (7) the effective linearity of a random permutation is:

$$EL(\text{random permutation}) = 1 + 8 \cdot 2^{-n} \cdot 1/8 \cdot 2^n = 2,$$

as asserted. ■

APPENDIX II
EFFECTIVE LINEARITY OF FEISTEL CONSTRUCTIONS

Let $f : \{0,1\}^{2n} \to \{0,1\}^{2n}$ be a 2-round Feistel construction with a permutation p_1 as the first round function, and a permutation p_2 as the second round function. Both permutations are defined over the space $\{0,1\}^n$.

An input difference (α_L, α_R) that enters this encryption scheme becomes after the first round (without the swap, which has no effect on our results) into $(\alpha_L \oplus \beta_1, \alpha_R)$, where $\alpha_R \xrightarrow{p_1} \beta_1$. After the second round the output difference is $(\alpha_L \oplus \beta_1, \alpha_R \oplus \beta_2)$ where $\alpha_L \oplus \beta_1 \xrightarrow{p_2} \beta_2$.

If the round functions are independent, the probability of the event $(\alpha_L, \alpha_R) \to (\alpha_L \oplus \beta_1, \alpha_R \oplus \beta_2)$ is $\Pr[\alpha_R \xrightarrow{p_1} \beta] \cdot \Pr[\alpha_L \oplus \beta_1 \xrightarrow{p_2} \beta_2]$. Thus, the difference distribution table of the 2-round Feistel construction contains in any entry the multiplication of the two related entries from p_1's and p_2's difference distribution tables.

[9] Up to the row corresponding to a zero input difference.

[10] Recall that in an XOR difference distribution table all values are even.

A. *Effective Linearity of a 2-Round Feistel Construction with Almost Perfect Nonlinear Permutations as Round Functions*

In this section we study the case where p_1 and p_2 are both APNPs.

Proposition 6: Let $f : \{0,1\}^{2n} \to \{0,1\}^{2n}$ be a 2-round Feistel construction with an APNP p_1 as the first round function, and an APNP p_2 as the second round function. Furthermore, assume that p_1 and p_2 are independent. Then $EL(f) = 7$.

Proof: If both p_1 and p_2 are APNPs, all the entries in the difference distribution table of the construction are either zero, 4, or 2^{n+1} (up to the $0 \to 0$ entry). For a given $\alpha = (\alpha_L, \alpha_R)$ such that $\alpha_R \neq 0$ there are two cases:

- If $\alpha_R \xrightarrow{p_1} \alpha_L$ through p_1, there is one entry in the row corresponding to α with the value 2^{n+1}, while all remaining non-zero entries are 4 and there are $(2^{2n} - 2^{n+1})/4$ such entries.
- If $\alpha_R \not\xrightarrow{p_1} \alpha_L$ through p_1, then all the non-zero entries are 4 and there are $2^{2n}/4$ such entries.

For any non-zero α_L there are about 2^{n-1} values of α_R for each of the above two cases. For $\alpha_L = 0$, all the values correspond to the second case. Thus, for the values of $\alpha = (\alpha_L, \alpha_R)$ there are $2^n(2^n - 1)/2 + (2^n - 1)$ rows with 2^{2n-2} entries of 4, and $2^n(2^n - 1)/2$ rows with one entry of 2^{n+1} and $2^{2n-2} - 2^{n-1}$ entries of 4.

The case when $\alpha_R = 0$ yields $2^n - 1$ rows in which there are 2^{n-1} entries of 2^{n+1} each. We note that the case $\alpha_L = \alpha_R = 0$ is not counted, as in this case $\alpha = 0$.

Therefore t_α has three possible values: $2^{2n-1} + 2^{2n-2} - 2^n$ for the first case, 2^{2n-2} for the second case, and 2^{3n-2} for the last case. Hence,

$$(2^{2n} - 1) \cdot \overline{t_\alpha} = \frac{2^n(2^n-1)}{2} \cdot (2^{2n-1} + 2^{2n-2} - 2^n) + \left(\frac{2^n(2^n-1)}{2} + (2^n - 1)\right) \cdot 2^{2n-2} + (2^n - 1) \cdot 2^{3n-2} \approx 2^{4n-1} + 2^{4n-2} \Rightarrow \overline{t_\alpha} \approx 2^{2n-1} + 2^{2n-2}.$$

Thus, for two Feistel rounds with APNPs as the round functions the effective linearity is close to $1 + 8 \cdot (2^{2n-1} + 2^{2n-2})/2^{2n} = 7$. ■

B. *Effective Linearity of a 2-Round Feistel Construction with Random Permutations as Round Functions*

The second case we study is when p_1 and p_2 are both random permutations.

Proposition 7: Let $f : \{0,1\}^{2n} \to \{0,1\}^{2n}$ be a 2-round Feistel construction with a random permutation p_1 as the first round function, and a random permutation p_2 as the second round function. Furthermore, assume that p_1 and p_2 are independent. Then $EL(f) = 14$.

Proof: Since p_1 and p_2 are random permutations, we may assume that the difference distribution tables of p_1 and p_2 can be treated according to a Poisson distribution, i.e., for a given entry of the difference distribution table of either of the permutations, $\Pr[DDT^p[\alpha, \beta] = 2k] = e^{-1/2}(1/2)^k/k!$. There are three scenarios with respect to $\alpha = (\alpha_L, \alpha_R)$:

- $0 \neq \alpha_R \xrightarrow{p_1} \alpha_L$ through p_1.
- $0 \neq \alpha_R \not\xrightarrow{p_1} \alpha_L$ through p_1.

Value	Ratio	Value	Ratio
0	$2e^{-1/2} - e^{-1} = 0.845$	4	$(1/2 \cdot e^{-1/2})^2) = 0.092$
8	0.046	12	0.008
16	0.007	20	0.0001
24	0.002	28	0.0000003

TABLE II

THE RATIO OF ENTRIES IN THE DIFFERENCE DISTRIBUTION TABLE WITH A GIVEN VALUE

- $\alpha_R = 0$.

The last case is the easiest to analyze, as the non-zero entries of the corresponding rows of the difference distribution table are 2^n times their value in the difference distribution table of p_2. For these values of α:

$$t_\alpha = \sum_{k=1}^{2^n-1} \frac{2^n e^{-1/2} \cdot 2^{-k}}{k!} \cdot \frac{(2^n \cdot k)\cdot(2^n \cdot k - 1)}{2} =$$

$$\frac{2^n}{2} \cdot \left[\frac{3}{4} \cdot 2^{2n} - \frac{1}{2} \cdot 2^n\right] = \frac{3}{8} \cdot 2^{3n} - \frac{1}{4} \cdot 2^{2n}$$

The second case is also easy to analyze, as each entry (α, β) where $\alpha = (\alpha_L, \alpha_R)$ and $\beta = (\beta_L, \beta_R)$ is a multiplication of the entries corresponding to $(\alpha_R, \beta_L \oplus \alpha_L)$ of p_1 and of $(\beta_L \oplus \alpha_L, \beta_R \oplus \alpha_L)$ of p_2. By the assumption, each of these entries has a known distribution, and we deduce that a fraction of $((1/2) \cdot e^{-1/2})^2$ of the entries of the difference distribution table of f have a value of 4, about $2 \cdot ((1/2) \cdot e^{-1/2}) \cdot (1/8) \cdot e^{-1/2})$ of the entries have a value of 8, etc. We give in Table II the expected ratio of entries with a given value. Taking all this information together allows to compute that for this kind of α the average value of t_α is $t_\alpha = 0.875 \cdot 2^{2n}$.

The first case can be split into two sub-cases. When the difference that enters p_2 is zero, the whole function behaves just like in the third case. When the difference that enters p_2 is non-zero, the whole function behaves just like in the second case. However, it is not clear how to find out which rows correspond to each case. In order to overcome this problem, we use the fact that actually we compute the average value of t_α over all the possible values of α. Hence, we can use symmetry arguments to claim that the average number of cases for which a zero difference enters p_2 is $2^n - 1$.

Taking all this information into consideration and using the average assumption, we have $2 \cdot (2^n - 1)$ entries that have $t_\alpha = (3/8) \cdot 2^{3n} - (1/4) \cdot 2^{2n}$, while the remaining entries have $t_\alpha = 0.875 \cdot 2^{2n}$. Thus,

$$\overline{t_\alpha} \approx \frac{2^{n+1} \cdot ((3/8) \cdot 2^{3n} - (1/4) \cdot 2^{2n}) + (2^{2n} - 2^{n+1}) \cdot 0.875 \cdot 2^{2n}}{2^{2n}} =$$

$$1.625 \cdot 2^{2n}.$$

We conclude that the effective linearity of a 2-round Feistel construction with random permutations as round functions is approximately $1 + 8 \cdot (13/8) \cdot 2^{2n} \cdot 2^{-2n} = 14$. ∎

C. Effective Linearity of a 2-Round Feistel Construction with Independent Round Functions

We now deal with the more general case, where the two round permutations p_1 and p_2 have no specific properties. However, we still assume that the round functions are independent. We prove that in this case, the effective linearity of f satisfies $EL(f) \geq EL(p_1)EL(p_2)$.

Proposition 8: Let $f : \{0,1\}^{2n} \to \{0,1\}^{2n}$ be a 2-round Feistel construction with a permutation p_1 as the first round function, and a permutation p_2 as the second round function. Furthermore, assume that p_1 and p_2 are independent. Then $EL(f) \geq EL(p_1)EL(p_2)$.

Proof: The proof is essentially similar to the proof of the case of random permutations as the round functions. We inspect the difference distribution table of the 2-round construction (from $2n$ bits to $2n$ bits) in order to compute the effective linearity of the construction. As in the previous examples there are three cases:

- $0 \neq \alpha_R \xrightarrow{p_1} \alpha_L$ through p_1.
- $0 \neq \alpha_R \overset{p_1}{\not\to} \alpha_L$ through p_1.
- $\alpha_R = 0$.

As before, the last case is the easiest to analyze, as the non-zero entries of the corresponding rows of the difference distribution table are 2^n times their value in the difference distribution table of p_2. Thus, the contribution of such rows to t_α is at least 2^{2n} times the contribution of the corresponding rows in the DDT of p_2 to $t^{p_2}_{\alpha_L}$. That is, for each such row we have $t_\alpha \geq 2^{2n} t^{p_2}_{\alpha_L}$. Note that while for the high values in the difference distribution table of p_2 the inequality is close to be an equality, for the lower values the inequality is sharper. The most apparent example are the entries in the DDT of p_2 whose value is 2. Such entries contribute 0 to $t^{p_2}_{\alpha_L}$, while the corresponding values in the DDT of f contribute $2^n(2^n - 1)$ to t_α.

The first case can be divided to the rows that behave like the second case and the rows that behave like the third case. Again, taking the average over all the possible α values, we can use symmetry arguments to claim that the number of rows that behave like the third case is $2^n - 1$ on average. These rows satisfy $t_\alpha \geq 2^{2n} t^{p_1}_{\alpha_R}$.

For the second case, we inspect the combined difference distribution table of f and find that:

$$t^f_\alpha = \sum_{\beta \in \{0,1\}^{2n}} \binom{DDT^f(\alpha,\beta)/2}{2} =$$

$$= \sum_{\substack{\beta_1, \beta_2 \in \{0,1\}^n, \\ \beta_1 \neq \alpha_L}} \binom{DDT^{p_1}(\alpha_R, \beta_1) DDT^{p_2}(\alpha_L \oplus \beta_1, \beta_2)/2}{2} \geq$$

$$\geq 8 \cdot \sum_{\substack{\beta_1 \in \{0,1\}^n, \\ \beta_1 \neq \alpha_L}} \binom{DDT^{p_1}(\alpha_R, \beta_1)/2}{2} \cdot$$

$$\sum_{\beta_2 \in \{0,1\}^n} \binom{DDT^{p_2}(\alpha_L, \beta_2)/2}{2} = 8 \cdot t^{p_1}_\alpha \cdot t^{p_2}_\alpha$$

Note that here we use the independence of the round functions and the fact that we count the average over all the possible values of α to assume that we can replace $DDT^{p_2}(\alpha_L \oplus \beta_1, \beta_2)$ by $DDT^{p_2}(\alpha_L, \beta_2)$ that is independent of β_1.

Therefore, for this construction:

$$\overline{t_\alpha} \geq \frac{2^n \cdot 2^{2n} \cdot \overline{t^{p_1}_\alpha} + 2^n \cdot 2^{2n} \cdot \overline{t^{p_2}_\alpha} + 2^{2n} \cdot 8 \cdot \overline{t^{p_1}_\alpha} \cdot \overline{t^{p_2}_\alpha}}{2^{2n}} \approx$$

$$2^n \cdot (\overline{t^{p_1}_\alpha} + \overline{t^{p_2}_\alpha}) + 8 \cdot \overline{t^{p_1}_\alpha} \cdot \overline{t^{p_2}_\alpha}$$

Thus, the effective linearity of the two round Feistel construction satisfies

$$EL(f) \geq 1 + 8 \left(2^n \cdot (\overline{t^{p_1}_\alpha} + \overline{t^{p_2}_\alpha}) + 8 \cdot \overline{t^{p_1}_\alpha} \cdot \overline{t^{p_2}_\alpha}\right) / 2^{2n} =$$

$$EL(p_1)\cdot EL(p_2) - 8\overline{t^{p_1}_\alpha}\cdot 2^{-n} - 8\overline{t^{p_2}_\alpha}\cdot 2^{-n} + 8\cdot 2^n\cdot(\overline{t^{p_1}_\alpha}+\overline{t^{p_2}_\alpha})/2^{-2n} = EL(p_1)\cdot EL(p_2)$$

This completes the proof of the proposition. ■

We note that there are constructions for which the above inequality is sharp. For example, for APNPs and random permutations the effective linearity of a 2-round Feistel structure is greater than the multiplication of the effective nonlinearities of the round functions. On the other hand, if both p_1 and p_2 are affine permutations then we have an equality. Hence, the inequality cannot be improved in the general case.

The same reasoning can be used when the round functions are non-bijective. In this case, as the round functions are not bijective, the number of zero entries in the difference distribution table is bigger, and there is a possibility that $p_1(x) = p_1(y)$ even if $x \neq y$. Obviously, this implies that the probability of the event $\alpha = \gamma$ is higher.

D. Effective Linearity of Reduced-Round DES

In this section we apply the general results established in the previous sections to approximate the effective linearity of reduced-round variants of DES [15].

Proposition 9: Denote by E_i an i-round reduced variant of DES. The effective linearities of $\{E_i\}$ satisfy:

1) $EL(E_1) = 320\cdot 2^{32} - 1$,
2) $EL(E_2) = 103039$,

Proof: For 1-round DES, the effective linearity can be computed explicitly using Equation 7. We tried all possible input differences to the F-function (the round function of a single DES round), and using Equation 7 we found out that the effective linearity of DES's F-function is 319. In the computation of the effective linearity of the full round, each pair of (input difference, output difference) to the F-function is actually repeated 2^32 times, once for each input difference in the left word of the plaintext (the one which does not enter the round function). Moreover, as we are dealing with a Feistel construction, it is possible to set the input difference of the right half to be 0 (as long as the left half is non-zero), and thus, by considering also these $2^{32}-1$ differentials with probability 1, we find out that the EL of 1-round DES is $319\cdot 2^{32} + 2^{32} - 1 = 320\cdot 2^{32} - 1$, as asserted.

In order to compute the effective linearity of 2-round DES we used the formula

$$EL(f) = -1 + 2^{-2n}\cdot \sum_{\alpha,\beta\in GF(2^n),\alpha\neq 0} [DDT^f(\alpha,\beta)]^2.$$

Denote the F-functions of the two rounds by p_1 and p_2. Since the effective linearity of the F-functions is 319, we get

$$2^{-64}\cdot \sum_{\alpha,\beta\in GF(2^{32}),\alpha\neq 0} [DDT^{p_i}(\alpha,\beta)]^2 = 320, \qquad (10)$$

for $i = 1, 2$.

As in the previous sections, the differentials of 2-round DES can be divided to three classes:

1) Differentials in which the input differences for both the F-functions are non-zero,
2) Differentials in which the input difference in the first round is non-zero and the input difference in the second round is zero,
3) Differentials in which the input difference in the first round is zero and the input difference in the second round is non-zero.

Recall that the input difference to the second round function is the XOR of the input difference in the left half of the plaintexts and the output difference of the first round function. As a result, for each differential of the F-function of the first round, and for each differential of the F-function of the second round, there exists a unique input difference for which the two 1-round differentials are combined into a 2-round differential. Hence, we have

$$2^{-128}\cdot \sum_{\alpha,\beta\in GF(2^{64}),\alpha\neq 0}[DDT^f(\alpha,\beta)]^2 =$$
$$2^{-128}\cdot \sum_{\alpha',\beta',\alpha'',\beta''\in GF(2^{32}),\alpha',\alpha''\neq 0}([DDT^{p_1}(\alpha',\beta')][DDT^{p_2}(\alpha'',\beta'')])^2$$
$$+2^{-128}\cdot \sum_{\alpha',\beta'\in GF(2^{32}),\alpha'\neq 0}[DDT^{p_1}(\alpha,\beta)]^2\cdot 2^{64} +$$
$$2^{-128}\cdot \sum_{\alpha'',\beta''\in GF(2^{32}),\alpha''\neq 0}[DDT^{p_2}(\alpha'',\beta'')]^2\cdot 2^{64},$$

where the first summand corresponds to the first class of differentials, the second summand corresponds to the second class of differentials and the last summand corresponds to the third class of differentials. Hence, using Equation 10 we get

$$2^{-128}\cdot \sum_{\alpha,\beta\in GF(2^{64}),\alpha\neq 0}[DDT^f(\alpha,\beta)]^2 = 320^2+320+320 = 103040,$$

and thus

$$EL(E_2) = -1 + 2^{-128}\cdot \sum_{\alpha,\beta\in GF(2^{64}),\alpha\neq 0}[DDT^f(\alpha,\beta)]^2 = 103039,$$

as asserted.

■

Appendix III
Effective Linearity of Reduced-Round AES

A. Effective Linearity of 1-Round AES

We start with computing the effective linearity of 1-round AES.

Proposition 10: The effective linearity of 1-round AES is $2^{25.54}$.

Proof: For the computation of the effective linearity of 1-round of AES we use the knowledge on the distribution of the differentials. We note that when considering 1-round differentials of AES it is sufficient to discuss only the S-box layer (ByteSub), as the remaining operations are linear, and thus do not affect the probability of the differentials.

The S-box layer is composed of 16 parallel applications of the same 8-bit to 8-bit S-box. Each row in the difference distribution table of the S-box that corresponds to a non-zero difference has one entry with a value of 4, and 126 entries with a value of 2.

Hence, for each S-box there exist 255 differentials with probability 2^{-6}. These differentials can be used to construct differentials with probability 2^{-6} of the entire first round by setting a non-zero input difference in the marked S-box and zero differences in the other S-boxes. As such differentials can be constructed for each and every S-box, there are a total of $16 \cdot 255$ differentials with probability 2^{-6}. In a similar manner, for any non-zero input difference for one S-box, there are 126 output differences with probability 2^{-7}. Using a similar construction we can generate $16 \cdot 255 \cdot 126$ differentials with probability 2^{-7}. We can continue to construct 1-round differentials with i active S-boxes of which in j S-boxes the probability of the differential is 2^{-6} while in the remaining $i-j$ S-boxes the probability is 2^{-7}.

Since a differential $\alpha \to \beta$ with probability p contributes p^2 to the effective linearity, we can find the effective linearity of 1-round AES according to Equation (7):

$$EL = -1+\sum_{i=1}^{16}\sum_{j=0}^{i} 255^i \cdot \begin{pmatrix} 16 \\ i \end{pmatrix} \cdot \begin{pmatrix} i \\ j \end{pmatrix} \cdot 126^{i-j} \cdot 2^{-2\cdot(7i-j)} = 2^{25.54},$$

as asserted. ■

B. Effective Linearity of 2-Round AES

Now we turn to compute the effective linearity of 2-round AES. Computing the effective linearity of 2-round AES is not an easy task. Unlike the case for 1-round AES, for 2-Round AES it is infeasible to count all the possible differentials. However, we can treat 2-round AES as a parallel application of four identical components composed of a layer of four S-boxes, MixColumns operation, XOR with a key, and another layer of S-boxes.[11] Then we can adopt the proof of Proposition 8 to get a lower bound on the effective linearity of 2-round AES:

Proposition 11: The effective linearity of 2-round AES is at least 49.

Proof: The proof of Proposition 8 can be easily adopted to permutations of the form $h(x, y) = (f(x), g(y))$. For such permutations $EL(h) \geq EL(f) \cdot EL(g)$.

In addition, it is possible to prove that the parallel application of two almost perfect nonlinear permutations achieves the minimal possible effective linearity for a parallel application construction. This is due to the joint difference distribution table, that contains only the minimal possible entries — 0, 4, and 2^{n+1} (up to the $0 \to 0$ entry)). Thus, the minimal effective linearity of two parallel components is at least as the minimal effective linearity of two APNPs, which is 7. Given the fact that two such double components are used in parallel, it follows that $EL(\text{2-Round AES}) \geq 7 \cdot 7 = 49$. ■

We experimentally computed the value of $EL(p)$, where p is the 32-bit permutation composed of a layer of four S-boxes, MixColumns operation, XOR with a key, and another layer of S-boxes. We found out that $EL(p) = 2$ (by enumerating all possible quartets). We also found out that the difference distribution table of this p is very similar to the difference distribution table obtained for a random permutation. As this is the case, we can conjecture that the parallel application of two components has an effective linearity of 14. The parallel application of two double components is thus expected to have effective linearity of at least 196.

We note that by treating the single component as a random permutation, it is possible to compute a closer estimation of the effective linearity of a 2-round AES. Using the same techniques as used in Appendix II-B, we computed the effective linearity of four parallel applications of random permutations to be 244.

Thus, assuming the difference distribution table of one component is close to the difference distribution table of a random permutation, we can conjecture that the effective linearity of 2-round AES is 244. Thus, we conclude this paper with a conjecture:

Conjecture 1: The effective linearity of 2-round AES is close to 244.

[11]There is no need to consider the ShiftRows operations, as they only mix the order of bytes.

ACKNOWLEDGMENTS

The authors would like to thank Osnat Ordan and Dana Cohen for their help in conducting the experiments, which verified our claims. It is also a pleasure to acknowledge the references and ideas expressed by Serge Vaudenay, Jennifer Seberry, Adi Shamir, and Eli Biham. We would also like to thank the anonymous referees for their valuable comments and insightful suggestions.

Orr Dunkelman Orr Dunkelman was born in Ramat Gan, Israel. He has acquired a B.A. degree in Computer Science (summa cum laude) in the excellence program of the Technion - Israel Institute of Technology, in 2000. He continued in a direct path to Ph.D. degree in the Technion, where he graduated at 2006.

Currently, Orr is a visiting fellow at the COSIC group, K.U. Leuven, in Belgium. Orr's research interests cover symmetric key design and analysis. His 9 years research in the field includes the first published attack on the full third generation block cipher KASUMI (A5/3), as well results on the AES, Serpent, IDEA, and the development of several new cryptanalytic techniques.

Dr. Orr Dunkelman is a member of International Association for Cryptologic Research (IACR), and is a co-founder of the Haifa Linux Club (Haifa's Linux User Group). During his Ph.D. studies, Orr was supported by the prestigious Clore scholarship. Currently Orr is supported by the Rothschild grant.

Nathan Keller Nathan Keller was born in Perm (Russia) in 1982, and immigrated to Israel in 1990. He achieved a B.A. degree in mathematics from the Technion (Haifa, Israel) in 1999, and an M.Sc. degree in mathematics from the Technion in 2002.

After serving in the Israeli Defense Forces in 2001-2004, he enrolled the Institute of Mathematics of the Hebrew University (Jerusalem, Israel) as a Ph.D. student. In addition, since May 2006 he works as a consultant in the Cryptography and Anti-Piracy Group in Microsoft Research (USA). His main topics of research are analytical methods in combinatorics and cryptanalysis.

Nathan Keller was a member of the Excellence Program of the Technion during his first degree studies (1998-1999), won the Zuchovitsky prize awarded by the Institute of Mathematics of the Hebrew University (2007), and is currently a recipient of the Adams Fellowship awarded by the Israeli Academy of Sciences and Humanities (2006-2009).